
Web-Based Free-Range Chicken Farm Productivity Prediction System Using Random Forest Algorithm

Wahyu Rezeky Abdillah Haris¹⁾, Hevlie Winda Nazry S²⁾

^{1,2)} Program Studi Sistem Informasi Fakultas Ilmu Komputer Dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara Medan

*Corresponding Author

Email: wrezeky64@gmail.com

Abstract

This research is motivated by the process of predicting native chicken egg production, which is still done manually based on farmer estimates, potentially producing inaccurate predictions and hampering production management efficiency. This study aims to develop a web-based native chicken farm productivity prediction system using the Random Forest algorithm. The study uses a quantitative approach with historical egg production data from one farm over six years. The dataset consists of 250 data with variables such as number of chickens, feed, temperature, humidity, age, and egg weight. Because it uses historical data, the study does not specify a population and survey sample. Data were collected through observation, interviews, and literature studies, then analyzed through data selection, preprocessing, data division 80% training and 20% testing, Random Forest Regression modeling, and evaluation using MAE, MSE, RMSE, and R². The results showed MAE 0.6702, MSE 1.2712, RMSE 1.1275, and R² 0.9621. In conclusion, the web-based system with Random Forest is able to produce egg number predictions with good performance and can be used to support evaluation and production planning for free-range chicken farms.

Keywords: *Free-Range Chicken, Machine Learning, Productivity Prediction, Random Forest, Web-Based System*

INTRODUCTION

Free-range chicken farming is a farming activity with economic potential through egg and meat production. Free-range chickens are a local Indonesian chicken breed derived from the red jungle fowl, which has undergone a domestication process and is able to adapt to its environment. This adaptability makes free-range chickens relatively resistant to disease and weather conditions compared to broiler chickens. Furthermore, free-range chickens are almost evenly distributed across Indonesia, despite their genetic diversity that causes differences in feather color, body size, and production capacity between individuals (Mutoffar et al., 2022). This situation indicates that free-range chicken farming has the potential to continue developing as a livestock business that produces food products such as eggs and meat. As its development progresses, increasing market demand also encourages free-range chicken farms to focus not only on maintenance aspects but also on improving their ability to manage production data so that farming activities can be carried out more effectively, efficiently, and measurably.

The productivity of free-range chicken farms, particularly egg production, is influenced by various rearing conditions. The number of chickens, feed amount, chicken age, coop temperature, and humidity are all variables related to egg production. These differences in conditions can cause egg productivity to fluctuate over time, requiring information that can more systematically describe farm conditions. On the farms studied, egg production is the primary output that needs to be monitored because monthly production is influenced by the number of chickens, feed, chicken age, and coop environmental conditions. The six-year historical production data used in this study also includes variables such as number of chickens, feed amount, chicken age, coop temperature, coop humidity, and egg production as a basis for describing the relationship between rearing conditions and egg production levels.

The problem identified is that the egg production prediction process is still carried out manually based on estimates and the experience of farmers. This condition often results in inaccurate predictions, which can impact production instability and reduce farm operational efficiency. This problem is further complicated by the fact that egg productivity is not solely related to a single factor but is influenced by several interrelated variables, such as feed, chicken age, temperature, and humidity. If data from these various variables is not systematically processed, the resulting information will be difficult to use as a basis for objectively estimating productivity levels. Therefore, an approach is needed that can process historical production data and various factors related to egg production to produce more measurable predictions.

The use of information technology and data processing is essential to address these issues. Production data, previously used only as records, can now be used as a source of information to develop productivity prediction models. This study used historical data on free-range chicken egg production collected over six years as the basis for model development. The data were obtained from production records on free-range chicken farms and used to illustrate the relationship between chicken rearing conditions and egg production. Systematic data processing is expected to help farmers obtain information on productivity levels based on available data, enabling farm condition evaluation and production planning to rely less on manual estimates.

One machine learning method that can be applied to this data processing is the Random Forest algorithm. Random Forest is a method that uses several decision trees built based on randomly selected data. Each tree produces a classification result, then the results from all trees are accumulated and selected based on the most frequently occurring result. The Random Forest structure consists of a root node, internal nodes, and leaf nodes, which are used in the decision tree formation process (Sandag et al., 2020). In its application, Random Forest also uses bootstrap sampling techniques and random feature selection at each branch to build several decision trees and produce more stable predictions. These capabilities make Random Forest suitable for both classification and prediction, as well as for data processing and reducing the risk of overfitting (Argadianata et al., 2025). Furthermore, Aini et al. (2024) demonstrated the use of Random Forest in the classification process, making this method relevant for application in prediction- and classification-based data processing.

The application of Random Forest in the prediction field has also been used in previous research. Prakoso et al. (2025) applied the Random Forest Regression algorithm to predict corn production and obtained an R^2 value of 0.98 on test data, while low MSE and MAE values indicate a high level of accuracy and prediction results close to the actual value. Nurdiyanto Yusuf's (2024) study on predicting beef production in Indonesia using Random Forest Regression also produced an R^2 value of 0.946, indicating the model's ability to explain approximately 94.6% of the variability in beef production data. Meanwhile, Priantama and Siswa's (2022) study showed that Correlation-Based Feature Selection optimization in the Random Forest Classifier can increase modeling accuracy from 91.66% to 97.22%. These results demonstrate that Random Forest has the potential to be used in predictive and classification data processing, including when the data contains multiple variables that need to be analyzed simultaneously.

Based on these problems, this study aims to develop a Web-Based Free-Range Chicken Farm Productivity Prediction System Using the Random Forest Algorithm that is able to process egg production data systematically and structured. The system was developed by utilizing the variables of number of chickens, amount of feed, chicken age, coop temperature, and humidity to generate egg productivity predictions. The application of Random Forest is aimed at producing more accurate productivity predictions compared to manual estimates and classifying productivity into low, medium, and high categories. The

urgency of this research lies in the need for farmers for data processing that can help evaluate farm conditions and support data-based decision making. The novelty of this research lies in the development of a web-based system that integrates historical data of free-range chicken egg production for six years with the Random Forest algorithm so that the prediction results can be used as material in production planning and more effective farm management.

RESEARCH METHODS

This research uses a quantitative method that refers to data processing based on quantity and size to produce empirical facts that can be explained systematically. The research implementation began with a literature study on the productivity of free-range chicken farms and a review of various journals related to the application of the Random Forest algorithm. This study served as the basis for determining the data processing approach and developing a productivity prediction system. Next, interviews were conducted with farm owners to obtain real-world information and data regarding the condition of free-range chicken farms. The data obtained were then used in the application of the Random Forest algorithm and the design of a web-based system to predict free-range chicken egg productivity. The selection of Random Forest was based on the algorithm's characteristics, which builds several decision trees through random selection of data and attributes, then combines the results of the decision trees to obtain predictions or classification results (Sandag et al., 2020). The application of this method is also supported by research by Prakoso et al. (2025) which showed that Random Forest Regression was able to produce production predictions with an R^2 value of 0.98, and research by Nurdiyanto Yusuf (2024) which obtained an R^2 value of 0.946 in predicting beef production.

Data collection was conducted through observation, interviews, and literature review. Observations were conducted directly on free-range chicken farms to obtain information on farm conditions and production activities, while interviews were conducted with the farm owner, Mr. Roby Haris, to obtain information on problems related to free-range chicken egg production. Literature review was conducted using guidebooks and journals related to free-range chicken farm productivity and the application of web-based Random Forest. The data analyzed were historical egg production data obtained from farmers, field observations, and monthly production records, with the variables number of chickens, amount of feed, chicken age, coop temperature, coop humidity, and number of eggs produced as target variables. Data analysis was carried out through the stages of data collection and selection, preprocessing, data division, application of Random Forest parameters, prediction process, and model evaluation. The dataset was divided into 80% training data and 20% testing data, then the Random Forest model was applied using the parameters `n_estimators`, `max_features`, `max_depth`, `min_samples_split`, `min_samples_leaf`, and `bootstrap`. Each decision tree generates a prediction, while the final result is determined by majority voting. Model performance is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), with smaller error values indicating better predictive performance. The use of attribute selection in the modeling process is also relevant to research by Priantama and Siswa (2022), which showed that Correlation-Based Feature Selection optimization can increase the accuracy of the Random Forest Classifier from 91.66% to 97.22%.

This study uses historical data on free-range chicken egg production collected over a six-year period from a single free-range chicken farm as the basis for developing a productivity prediction model. The dataset was obtained through farm production records and supported by field observations and information from farmers. The variables include the number of chickens, feed intake, age, coop temperature, coop humidity, egg production, and

egg weight. Productivity is used as a classification result. These data are used to illustrate the relationship between free-range chicken rearing conditions and egg production levels. This study does not define a population and sample in the form of a survey study and does not use statistical calculations such as Slovin or Cochran because the unit of analysis is historical production data from a single farm over six years. The use of historical data as the basis for modeling is in line with the research of Prakoso et al. (2025) who utilized statistical data to predict corn production using Random Forest Regression and the research of Nurdiyanto Yusuf (2024) who used production data to predict beef production in Indonesia.

The research procedure began with the collection of historical data on free-range chicken egg production obtained from farmer records, direct field observations, and monthly production records. The data was then selected by selecting variables related to egg production and eliminating irrelevant or redundant data to improve the quality of the prediction model. The next stage was preprocessing through data cleaning by removing duplicate data and handling missing values. The dataset was then divided into 80% training data for model building and 20% testing data for model performance testing. Next, Random Forest was applied with the parameters `n_estimators`, `max_features`, `max_depth`, `min_samples_split`, `min_samples_leaf`, and `bootstrap`. The Random Forest algorithm builds several decision trees based on randomly selected training data using the bootstrap sampling technique. Then, the prediction results from each tree are used to obtain the final result through a process of combining the decision tree results (Sandag et al., 2020). Each tree produces a prediction of egg production categories and the final result is determined by majority voting. While model evaluation was carried out using MAE, MSE, and RMSE to determine the closeness of the prediction results to the actual value. Once the model is created and evaluated, Random Forest is integrated into a web-based system so users can input laying hen data and obtain predicted egg production results as a basis for decision-making. The prediction results are then analyzed to provide information related to feed optimization, cage management, and increasing egg production in free-range chickens. The application of Random Forest in the prediction process is supported by previous research that demonstrates the method's ability to produce predictions with a high level of performance, such as research by Prakoso et al. (2025) and Nurdiyanto Yusuf (2024), as well as the application of Random Forest in classification by Aini et al. (2024) and Argadianata et al. (2025).

RESULTS AND DISCUSSION

Data Collection Results

The dataset used consists of 250 data points collected as the basis for processing and developing a chicken productivity prediction model. Each data point contains information on chicken rearing conditions and production results, which are used as variables in the study.

The variables in the dataset include the number of chickens, feed, temperature, humidity, age, number of eggs, and egg weight. These variables are used to describe the rearing conditions and productivity of the chickens in each observation. The data is then used in the preprocessing stage to ensure the data used in the modeling process is in the appropriate condition.

With a total of 250 data points, the dataset was then preprocessed and used as the basis for developing a predictive model. The processed data were then used to determine the model's ability to predict chicken productivity based on the available variables.

Data Processing Results and Prediction Modeling

Data Selection

The data selection stage involves selecting relevant attributes for use in predicting native chicken productivity. At this stage, attributes unnecessary for the modeling process are removed, resulting in only variables that contribute to the prediction process.

The Number attribute was removed because it only serves as an identifier or marker for each data item and does not provide information related to chicken productivity. Furthermore, the Number of Eggs attribute was not used as an input variable because it was designated as the target variable to be predicted by the model. Therefore, this attribute was separated from the predictor variables to prevent the target from being used as input in the modeling process.

After the selection process, the remaining data consists of attributes used as predictor variables to predict the number of eggs, an indicator of native chicken productivity. The results of this data selection are then used in the preprocessing stage before developing the predictive model.

Data Preprocessing

The data preprocessing stage is the process of preparing data before it is used in developing a chicken productivity prediction model. This stage aims to ensure that the data used is of good quality, complete, and meets the modeling requirements. Data that has passed the selection stage is then examined and processed through several preprocessing steps.

The data preprocessing stage in this study began with data cleaning. This process was carried out to ensure there were no issues in the data that could impact the modeling process. Data cleaning included checking for missing values and duplicate data.

Data Cleaning

Data cleaning is performed to ensure the selected data is clean and ready for use in the next processing stage. This process examines all 250 data points for missing values and duplicates.

Handling Missing Values

A blank value check was performed to determine whether any attributes in the dataset had missing values. Based on the results of an examination of 250 data sets, no missing values were found in the attributes used in the study. Therefore, no data entry or deletion was required due to missing values, allowing all data to be retained.

Duplicate Data Handling

A duplicate data check was performed to determine if there were any similar data items recorded more than once in the dataset. The results of the 250 data sets showed no duplicates, or zero. Therefore, no data deletion was necessary, and all 250 data sets were retained for subsequent processing.

Data Sharing

Data splitting was performed to separate the dataset into data used in the model training process and data used to test the model's predictive ability. In this study, the data split used an 80:20 ratio, with 80% for training data and 20% for testing data. Of the 250 data sets, 200 were used for training and 50 for testing.

Random Forest Modeling

After the data is divided into training and testing data, the next step is to apply the Random Forest Regression algorithm to build a predictive model for chicken productivity. The Random Forest Regression algorithm is used to study the relationship between input variables related to chicken rearing conditions and the target variable, egg count. The model is built using several decision trees, whose predictions are combined to obtain the final predicted value.

In this study, a Random Forest Regression model was built with several predetermined parameters, namely `n_estimators=20`, `max_depth=5`, `min_samples_split=3`, and `random_state=42`. These parameter settings are used to regulate the decision tree formation process so that the model can learn from the data in a controlled manner.

The `n_estimators=20` parameter is used to determine the number of decision trees formed in the Random Forest model, which is 20. Each tree generates a prediction based on the data used in the training process, and then the results from all trees are combined to produce the final prediction.

The `max_depth=5` parameter is used to specify the maximum depth of each decision tree, which is 5 levels. This depth limitation prevents the tree structure from becoming too complex, thus helping to reduce the risk of overfitting the training data.

The `min_samples_split=3` parameter specifies the minimum number of samples a node must contain before it can be split into subsequent nodes. In this study, a node must have at least 3 data points to be able to split.

Next, the `random_state=42` parameter is used to set a random seed so that the model building process can produce consistent results when run again using the same data and parameters. Therefore, this parameter configuration serves as the basis for training the Random Forest Regression model in this study.

Model Evaluation

Model evaluation was conducted to determine the ability of the Random Forest Regression algorithm to predict chicken productivity based on test data. The evaluation was conducted by comparing the model's predicted values with the actual values in the test data. In this study, the model evaluation used four metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2).

Based on the test results using the testing data, the MAE value was 0.6702, the MSE was 1.2712, the RMSE was 1.1275, and the R^2 was 0.9621. The MAE value of 0.6702 indicates that the average absolute difference between the model's prediction results and the actual number of eggs is approximately 0.67 units. Meanwhile, the MSE value of 1.2712 indicates the average squared error in the predictions produced by the model.

The RMSE value obtained, 1.1275, indicates the magnitude of the prediction error in the same units as the target variable, namely the number of eggs. The smaller the MAE, MSE, and RMSE values, the smaller the error generated by the model. Based on the obtained values, the model produces a relatively low level of prediction error.

Furthermore, the R^2 value of 0.9621 indicates that the model is able to explain approximately 96.21% of the variation in egg count based on the input variables used in the study. Meanwhile, approximately 3.79% of the remaining variation is beyond the model's capabilities and may be influenced by other factors not used in the study. Overall, the evaluation results indicate that the Random Forest Regression model has good predictive ability in estimating egg count as an indicator of chicken productivity.

Program Implementation

Program implementation is the stage of applying the results of the program design into an application that can be used.

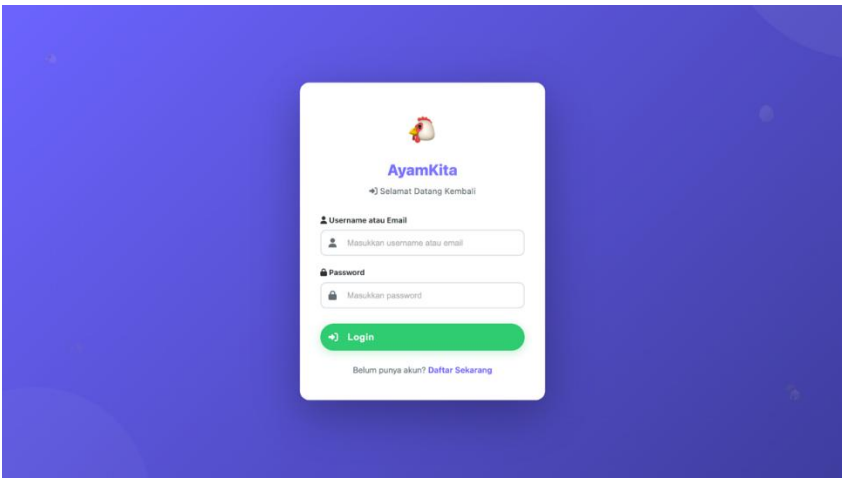


Figure 1. Login

Figure 1. is the admin login page, if the user fills in the username and password correctly, it will appear as in Figure 2.

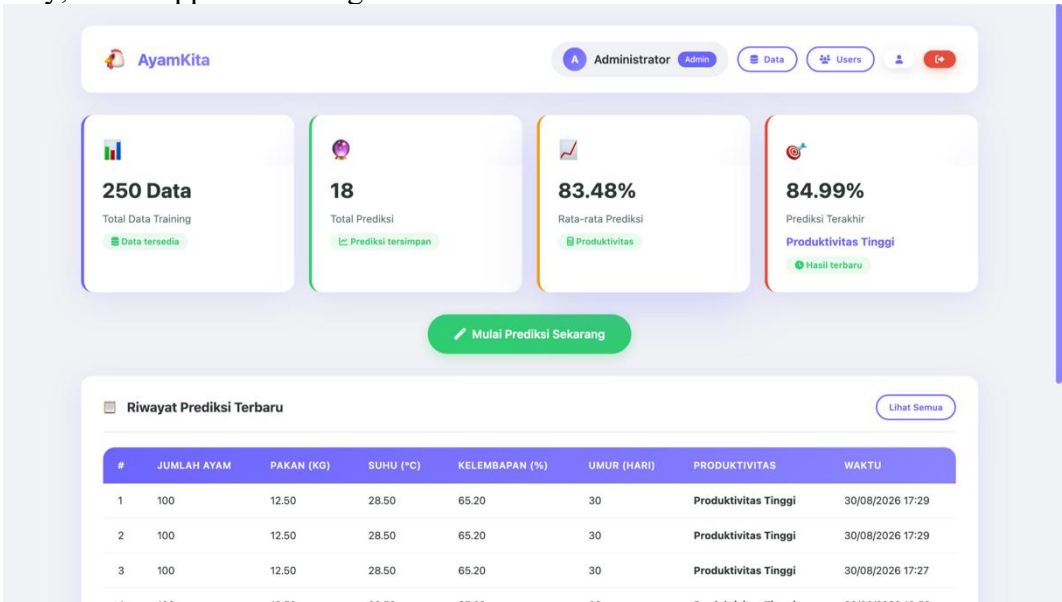


Figure 2. Dashboard

Figure 2 shows the dashboard page, which contains several buttons: data, users, and start prediction. If the user clicks the data button, the screen will appear as in Figure 3.

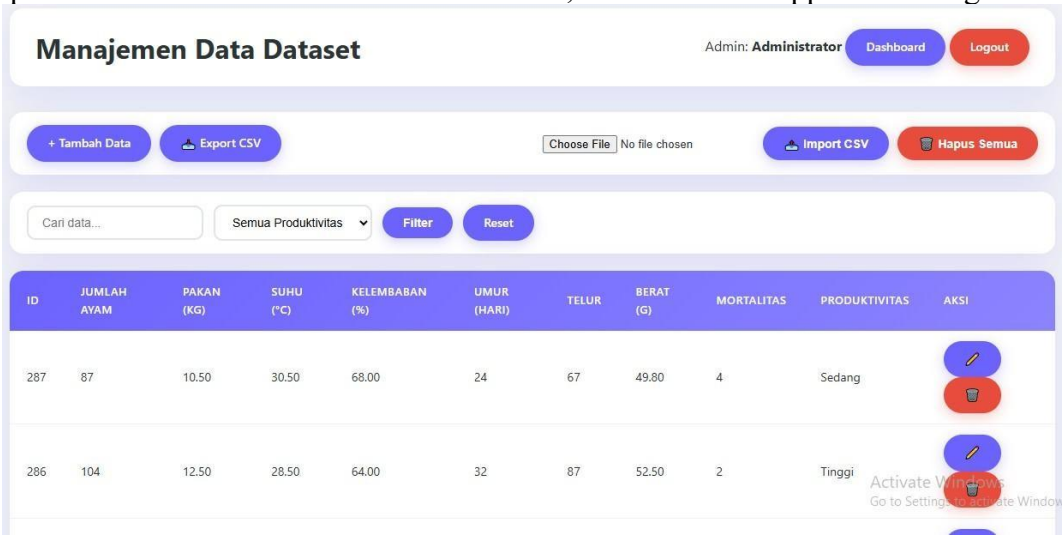


Figure 3. Data

Figure 3 is the admin data page, there are several buttons namely export csv, import csv, delete all, filter, reset, add data, edit data, and delete data. If export csv will download the csv file, if import csv will appear the csv file selection window, if click delete all then all data is deleted, if click filter will display the data according to the selected filter, if click reset will reset the filter results, if click add data will display the data adder, if click edit on one of the data will display the data editing form, and if click delete data on one of the data will delete one of the data. If click dashboard then click users it will appear as Figure 4.

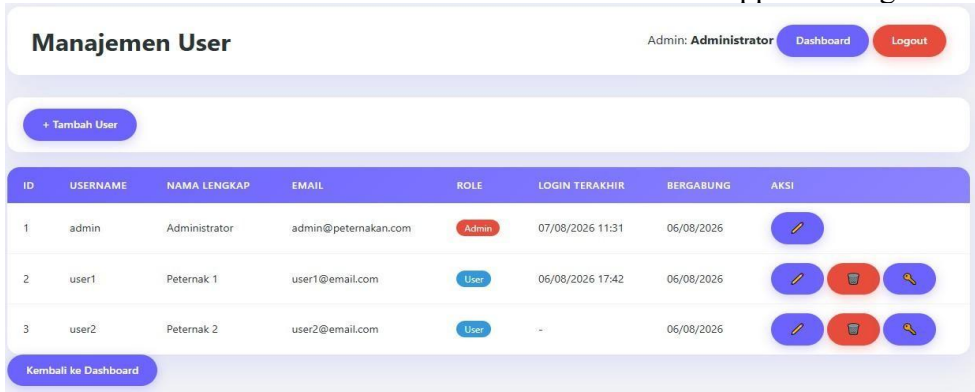


Figure 4. User

Figure 4 is the users page, with several buttons: add user, edit, and delete. Clicking add user will display a data entry form. Clicking edit on a record will display a data editing form. Clicking delete on a record will delete the selected record. Clicking account information will display Figure 5.

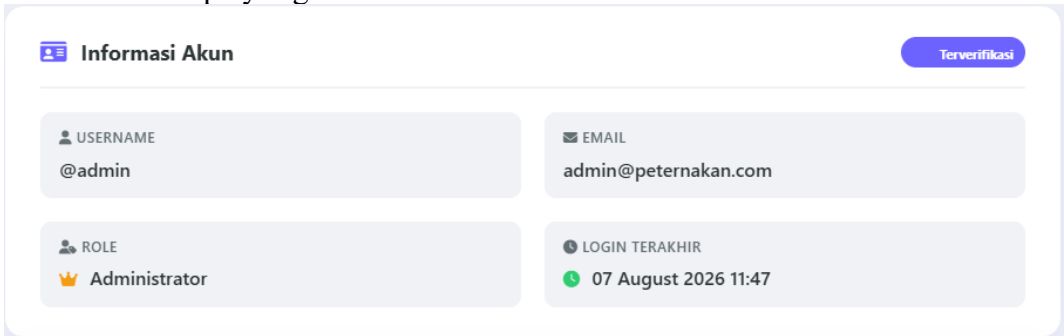


Figure 5. Account Information

If you click the dashboard button and click start prediction now, it will appear as in Figure 6.

Form Prediksi

Isi semua field

Jumlah Ayam *

100

→ Range: 50-150 ekor

Pakan *

12.5

→ Range: 5-20 kg

Suhu Kandang *

28.5

→ Optimal: 25-30 °C

Kelembaban *

65

→ Optimal: 60-70%

Umur Ayam *

30

→ Produktif: 20-40 hari

Prediksi Sekarang

Reset

Figure 6. Prediction

Figure 6 shows the prediction page, which has two buttons: "Predict Now" and "Reset." If the user enters all the data and clicks "Predict Now," the prediction results will appear. Clicking "Reset" will erase the entered data.

If the user wants to use the user/breeder section page, they must log out by clicking the logout button and then log in as in Figure 7.

AyamKita

Selamat Datang Kembali

Username atau Email

user1

Password

Login

Belum punya akun? [Daftar Sekarang](#)

Figure 7. User Login

Figure 7 is the user login page, if the user fills in the username and password correctly, it will appear as in Figure 8.

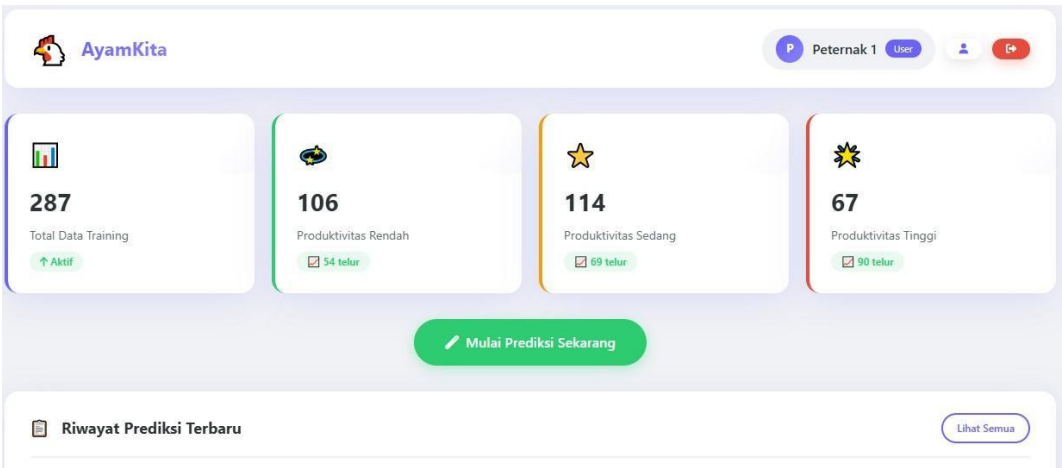


Figure 8. Dashboard

Figure 8 shows the dashboard page, with several buttons: "Users" and "Start Prediction." If the user clicks the "Account Information" button, the screen will appear as shown in Figure 9.

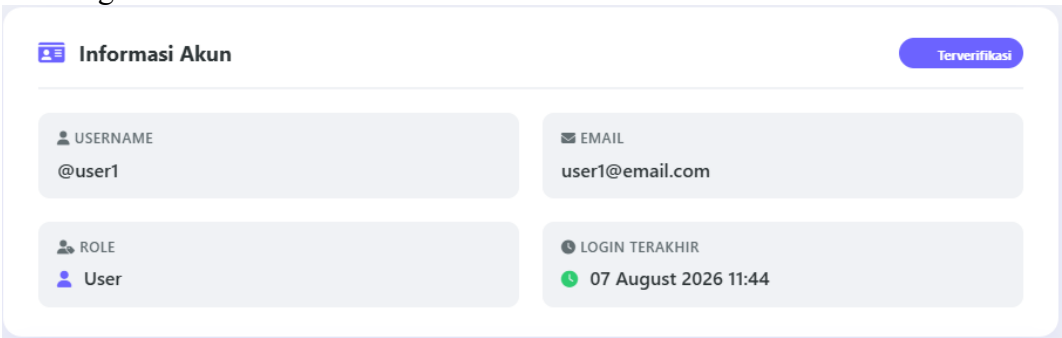


Figure 9. User Account Information

Figure 9 shows the account information page, which displays the logged-in user. If the user clicks the dashboard button and then the "Start Prediction Now" button, the display will appear as shown in Figure 10.

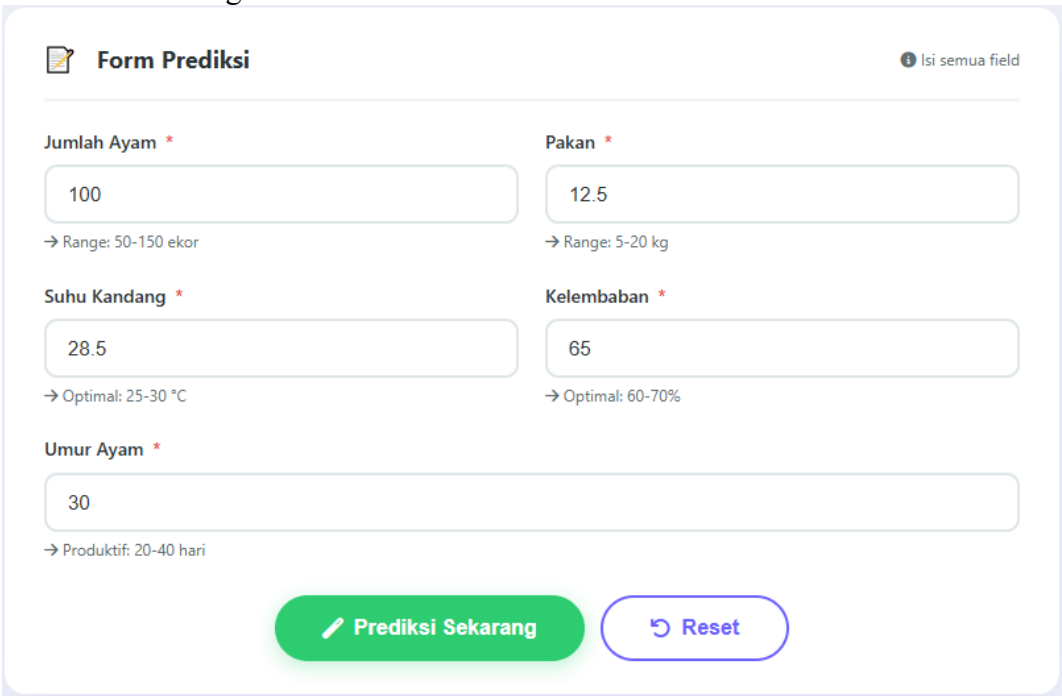


Figure 10. Prediction Form

Figure 10 shows the prediction page for entering test data for the random forest algorithm. Once all the data is filled in and you click Predict Now, the results will appear as in Figure 11.

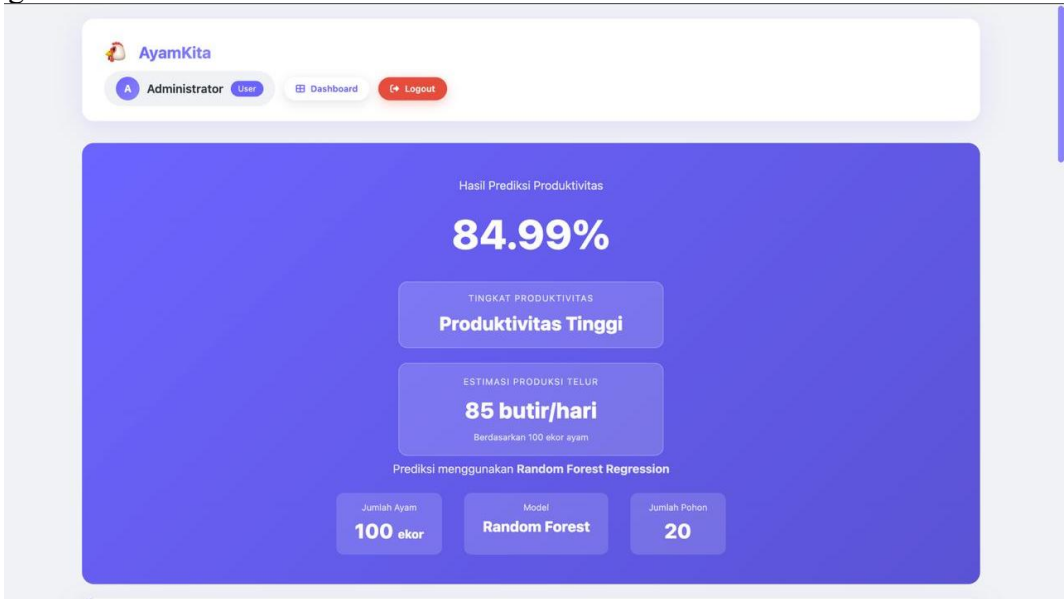


Figure 11. RF Algorithm Test Results

Program Testing

Black box testing is performed to determine whether each function in the system is operating according to functional requirements. Testing is performed without considering the program code structure, but rather focuses on the input, processing, and output generated by the system. The test results indicate that all key features of the system are operating as expected.

Table 1. Black Box Testing of Admin Login Page

No	Scenario Testing	Input	The results that It is expected	Results Testing	Conclusion
1	Login usingcorrect username and password	Valid username and password	The system displays the dashboard pageadmin	As expected	Succeed
2	Login usingusername or password that Wrong	Username/passwordinvalid	The system displayslogin failed message	As expected	Succeed
3	Login without filling in data	Empty	The system asks the user to fill in the datalogin	As expected	Succeed

Table 2. Black Box Testing of Admin Dashboard

No	Scenario Testing	Input	The results that It is expected	Results Testing	Conclusion
1	Open the Data menu	Click Data button	System displayData page	As expected	Succeed
2	Open the Users menu	Click the button Users	The system displays Users page	As expected	Succeed
3	Open the Start Prediction menu	Click the Start button Prediction	The system displays the page Prediction	As expected	Succeed

Table 3. Black Box Testing of Data Pages

No	Scenario Testing	Input	The results that It is expected	Results Testing	Conclusion
1	Export data	Click Export CSV	CSV file downloaded successfully	As expected	Succeed
2	Import data	Select CSV file	Data successfulentered into database	As expected	Succeed
3	Add data	Fill out the entire form	Data saved successfully	As expected	Succeed
4	Edit data	Changing data	Data successful updated	As expected	Succeed
5	Delete data	Click the Delete button	Data deleted successfully	As expected	Succeed
6	Delete all data	Click Delete All	All data was successfully deleted	As expected	Succeed
7	Data filter	Selecting a filter	Data is displayed according to the filter	As expected	Succeed
8	Reset filter	Click Reset	All data displayed again	As expected	Succeed

CONCLUSION

This research produces a web-based free-range chicken farm productivity prediction system by applying the Random Forest Regression algorithm. The system was built using 250 historical egg production data with variables of number of chickens, feed, temperature, humidity, age, and egg weight. The model uses a data split of 80% for training and 20% for testing with parameters `n_estimators=20`, `max_depth=5`, `min_samples_split=3`, and `random_state=42`. The evaluation results on the testing data show a MAE value of 0.6702, MSE of 1.2712, RMSE of 1.1275, and R^2 of 0.9621. These findings indicate that the model is able to provide egg number predictions with a relatively low error rate and explains 96.21% of the variation in egg number based on the input variables used. Practically, the web-based system can help farmers process production data more systematically and obtain predictive information to support farm condition evaluation and production planning.

However, this study has limitations because the dataset used was only from a single farm and consisted of 250 data points, so the model's generalizability to different farming conditions is still limited. The variables used also did not cover all factors that could potentially influence productivity, such as feed type, chicken health, cage density, lighting, vaccination, and disease history. Therefore, further research is recommended to increase the number and diversity of data, expand the research object, and include additional relevant variables. Comparison of Random Forest with other algorithms such as SVM, Decision Tree, KNN, Naïve Bayes, or XGBoost can also be conducted to determine the best-performing method. Further system development can be directed at a mobile platform and the addition of action recommendation features based on prediction results so that the system not only provides productivity information but also supports decision-making and corrective actions by farmers.

REFERENCES

- Aini, N., Arif, M., Agustin, I. T., & Toyibah, Z. B. (2024). Implementasi algoritma Random Forest untuk klasifikasi bidang MSIB di Prodi Pendidikan Informatika. *Jurnal Informatika*, 11(1), 11–15. <https://doi.org/10.31294/inf.v11i1.20637>
- Argadinata, A. P., Fatah, D. A., & Sukri, H. (2025). Klasifikasi kualitas buah apel menggunakan metode Random Forest. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 2016–2022. <https://doi.org/10.36040/jati.v9i2.12854>
- Mutofar, M. M., Naseer, M., & Fadillah, A. (2022). Klasifikasi kualitas air sumur menggunakan algoritma Random Forest. *NARATIF: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, 4(2), 138–146. <https://doi.org/10.53580/naratif.v4i2.160>
- Nurdiyanto Yusuf. (2024). Prediksi produksi daging sapi di Indonesia menggunakan Random Forest Regression: Analisis data 2018–2025. *Jurnal Ilmiah Teknik*, 3(2), 134–142. <https://doi.org/10.56127/juit.v3i2.1620>
- Prakoso, A. B., Putra, M. A., Hilmi, M. H., Risky, S. Y. P., & Maulindar, J. (2025). Penerapan algoritma Regresi Random Forest untuk prediksi produksi jagung menggunakan data statistik sistem pertanian cerdas Smart City. *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB) 2025*. <https://doi.org/10.47701/19h5ny78>
- Priantama, Y., & Siswa, T. A. Y. (2022). Optimasi Correlation-Based Feature Selection untuk perbaikan akurasi Random Forest Classifier dalam prediksi performa akademik mahasiswa. *JIKO (Jurnal Informatika dan Komputer)*, 6(2), 251–260. <https://doi.org/10.26798/jiko.v6i2.651>
- Sandag, G. A. (2020). Prediksi rating aplikasi App Store menggunakan algoritma Random Forest. *CogITO Smart Journal*, 6(2), 167–178. <https://doi.org/10.31154/cogito.v6i2.270.167-178>